

U.S. DEPARTMENT OF COMMERCE  
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION  
NATIONAL WEATHER SERVICE  
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 432

NCEP/EMC OBSERVING SYSTEM EXPERIMENT (OSE) ON THE EFFECT OF  
LOSS OF RUSSIAN RAOB DATA

M. STEVEN TRACTON, ROBERT KISTLER AND YUEJIAN ZHU  
National Centers for Environmental Prediction/Environmental Modeling Center

AND

JIM GIRAYTYS  
WMO Contractor

MAY 2001

THIS IS AN UNREVIEWED MANUSCRIPT, PRIMARILY INTENDED FOR INFORMAL  
EXCHANGE OF INFORMATION AMONG NCEP STAFF MEMBERS.

# NCEP/EMC OBSERVING SYSTEM EXPERIMENT (OSE) ON THE EFFECT OF LOSS OF RUSSIAN RAOB DATA

by

M. Steven Tracton, Robert Kistler, and Yuejian Zhu  
National Centers for Environmental Prediction /Environmental Modeling Center  
NOAA/NWS, Washington, DC  
and  
Jim Giraytys  
WMO Contractor

## 1. Background

The number of sites in the Russian Federation taking upper air observations using radiosondes (RAOBs) decreased by 65% to 70% from January 1994 to December 1999, largely as a function of budget restrictions. Such a decrease might be expected to affect the skill of Numerical Weather Prediction forecasts. To address this issue, the Open Program Area Group on Data Processing and Forecast Systems of the WMO formed an Expert Team whose mission was to assess the possibility of using routine verification statistics as an alternative to conducting the more costly and time consuming Observing System Experiments (OSEs). The expert Team adopted the following working hypothesis:

*“It is possible to establish meaningful impact in the radiosonde network in Russia through evaluation of readily available verification scores of global and/or regional scale operational forecast models”.*

The results of this investigation are documented in the WMO publication, *“Study of the Impact of the Loss of Russian Federation RAOBS on NWP Verification Statistics in the Northern Hemisphere”* (available at <http://www.wmo.ch/web/www/reports.html>). In summary, those results neither proved nor disproved the hypothesis. No clear signal in the verification statistics was found, except *possibly* some degradation in skill over Asia and North America. Given changes in NWP systems over the years and likely natural variations in predictability (e.g., associated with circulation regime, seasonal trends), it is very difficult to detect a real signal and equally difficult to assign any change to a particular cause. One cannot conclude from this study that the loss of Russian Federation RAOBs does result in a meaningful loss of skill in NWP, only that the approach used in the investigation was not adequate to uncover a signal in the noise of the several factors which might lead to a change in routine verification scores.

To investigate further whether loss of the Russian data does have a meaningful impact on NWP skill, NCEP agreed to conduct an OSE using its Reanalysis system and data set. A brief description of the experiments is provided in Section 2. Illustration of results is provided in

Section 3, and a summary and discussion concludes the report (Section 4).

As will be shown, there is some systematic loss of skill in forecasts regionally (Asia, Alaska and Northern Canada) at short ranges (<3days). In the medium range (3-8 days) results are entirely consistent with the earlier investigation - no clear signal of a degradation in NWP forecast skill resulting from loss of the Russian RAOBs. In the OSE loss of the data unequivocally does produce an impact on forecasts, but beyond a few days that impact is largely random (some positive, some negative) and small relative to the total error in forecasts. The most important caveat to these results is that they apply only to the particular NWP system used and period of study selected.

## 2. Experimental Design

The OSE was run using the NCEP Reanalysis data assimilation system and data set. The system includes the NCEP global spectral model operational ("MRF") in 1995 with T62 (~210 km) horizontal resolution and a three-dimensional variational (3DVAR) analysis scheme. The data set includes RAOBs, TOVS temperature soundings, cloud-tracked winds, aircraft observations, land and ocean surface reports, etc. Further details can be found in Kistler, et al. (February, 2001 BAMS). The experiment period selected was January, 1994, a month where subjectively (based on appraisal of circulation patterns and weather systems) it appeared loss of the RUSSIAN RAOBs could have a notable impact upon NWP forecasts.

Three experiments were run:

EXP1: the control - all available observations

EXP 2: all observations, **except** those Russian RAOBS available in January, 1994, but NOT available in January, 2000 (nearly identical to the comparison between January, 1994 and December 1999 shown in Figs. 1 and 2, respectively, in the WMO report)

EXP 3: all RUSSIAN RAOBS removed

EXP3 was run as a baseline sensitivity test to assess the effect of removing all Russian Federation radiosonde reports. The relevant data sets were assimilated starting Dec. 1, 1993 and continued through Jan. 31, 1994 (December provided a one month "spinup"). Forecasts to 8 days were run from the 00Z analyses of each experiment for each day of January, 1994. Verifications were in the form of anomaly correlation (AC) scores and RMSE of forecasts with respect to the control set of analyses over the Northern Hemisphere and selected subregions thereof. Subjective case study evaluation was also performed to confirm and complement these objective verifications. Also, the subjective appraisal suggests that additional objective verifications against radiosonde observations likely would not change the results.

### 3. Results

#### 3.1 Short Range Forecasts

Verification statistics indicate that at shorter ranges a small, but systematic negative impact is felt in the Asian and JMA (centered about Japan) domains. This is demonstrated by the RMSE verifications for 2-day forecasts shown in Fig. 1 and 2. In these and the following figures black is for EXP1, the red for EXP2, and green for EXP3. The means over all cases is displayed on the bottom right of the top figure. The verification date (not initial time) is shown on the abscissa. The bottom portion of the figure displays the time series of the difference between EXP1 and EXP2.

The E1 errors on average over the period (20.25/25.06 for JMA/Asia) are slightly smaller than for EXP2 (20.89/25.72). More importantly, as seen clearly in the plots of EXP1-EXP2 scores, the negative impact occurs in most cases. Meteorological significance from subjective evaluation, however, is dubious. Over Alaska and Northern Canada, where there are no separate objective verifications, it appears qualitatively that forecast differences are larger than over Asia and the JMA regions. At day 2, and to a limited extent at day 3, there is a tendency generally for the EXP2 to have larger errors than EXP1. As an example, Fig. 3 shows for a particular case the difference between EXP1 and EXP2 2-day 500 mb height forecasts. In comparison to the corresponding EXP1 and EXP2 error fields (Fig. 4), one can observe that areas of forecast difference are reflected mostly as larger errors in EXP2. Note too the emergence of errors exclusive of the impact of loss of Russian data. This is evident from the observation that in regions of EXP1/EXP2 differences the errors are generally larger than those differences and more clearly from the errors occurring outside the areas of EXP1/EXP2 differences.

Experiments with regional models would have to be performed to assess more thoroughly the significance of the loss of data on short term forecasting over regions in relatively close proximity to the differences in analyses due to the decline in Russian RAOBs. In that context, before serious non linearities develop and other sources of error become dominant, the sense of the analysis differences (presumed to be negative) is more likely to be felt. Finally here, note from EXP3 that the impact that would occur from removal of all Russian RAOBs is considerably larger, as seen objectively in Figs. 1 and 2 and observed also in subjective evaluation. This is commented upon further in Section 4.

#### 3.2 Medium Range Forecasts

Figure 5 shows the time series of AC scores of 5-day forecasts computed for the Northern Hemisphere north of 20°. It can be seen that there is a mix of positive and negative impacts with the difference in the mean between EXP1 and EXP2 very small (E2 actually better) and certainly not statistically significant. The same conclusion is drawn from the corresponding set of RMS errors (not shown). And, the same result applies to each of the verification sub regions shown in Figs. 6 - 8 and for the most part to day 3 forecasts as well (not shown).

The lack of significant effect on verification scores does not mean that the loss of the Russian RAOBS does not have an impact on the forecasts. This can be seen from the initial and 5-day EXP2-EXP1 differences shown in Fig. 9 for the same case discussed above. The relatively small differences that exist between analyses (day 0) in the polar regions over and to the north of Siberia evolve (via translation and downstream amplification/propagation) through those shown in Fig. 3 (day 2) to seemingly very significant levels both in magnitude and areal coverage. But, beyond 2-3 days those differences are generally small compared to the forecast error, whether it

be for EXP1 or EXP2 (Fig. 10). In effect, the signal of forecast impact due to loss of the Russian RAOBs is essentially noise in the context of the total forecast error. Or in the vernacular of NWP, the forecasts are much more alike than either is to the real atmosphere. From one case to the next or within one case from region to region subjective appraisal shows that, when it is possible to judge that one forecast is better than the other in some respect, there is no systematic preference for EXP1 or EXP2. By no later than days 2- 3, non-linear interactions dominate and the sense of forecast differences is essentially random.

#### 4. Summary and Discussion

The OSE was designed explicitly to isolate the effect of the loss of Russian RAOBs. Unlike the earlier study based on evaluation of routine verification scores, the impact is not obscured by before and after differences in atmospheric predictability (e.g., related to circulation regime), changes in the global observing system, or differences between models and data assimilation schemes. It was a perfectly "clean", *but limited*, experiment.

The results indicated a small, but systematic loss of skill at short ranges (< 3days) in the regions most local to the areas affected by the loss of data (Asia, JMA, Alaska and Northern Canada). In the medium range (3-8 days), results were remarkably comparable to those arrived at by the Expert Team - no degradation in forecast skill as a result of deterioration in the Russian Federation RAOB network. Beyond day 3, the loss of data does have a notable impact, but the differences between the with and without experiments (EXP1 and EXP2) are small and not systematic relative to the total error of either EXP1 or EXP2 predictions. The loss in skill associated with loss of the RAOBs over Russia is essentially noise in the context of other sources of error, i.e., analysis errors exclusive of the Russian data and inadequacies in the data assimilation system and forecast model.

It is worthwhile to note that the objective verifications and subjective evaluation of EXP3 does indicate a fairly large and more or less systematic degradation in forecast skill even at medium ranges as a result of removing all Russian Federation RAOBS. In most instances the result is making a bad forecast even worse. But in some, the loss of data clearly renders a reasonably good prediction less useful. In the context of this experiment, by chance or design, it appears the decrease, but not elimination, in Russian RAOBS was such as to minimize the impact on global NWP.

Of course, the principal caveat of this OSE is that it applies only to the particular numerical and analysis forecast system used (NCEP, 1995 vintage) and only for the particular period selected. To adequately generalize would require using the latest state-of-the art system (more than one) applied to several independent sample periods. That would be exceedingly costly in human and computer resources (and beyond anything currently planned at NCEP). And, if later studies, in fact, did demonstrate a clear and meteorologically significant loss in skill from the decline in Russian RAOBs, one likely would have to address the cost effectiveness of alternative observing systems (e.g., aircraft ascent/decent soundings) in comparison to reconstituting the Russian RAOB network.

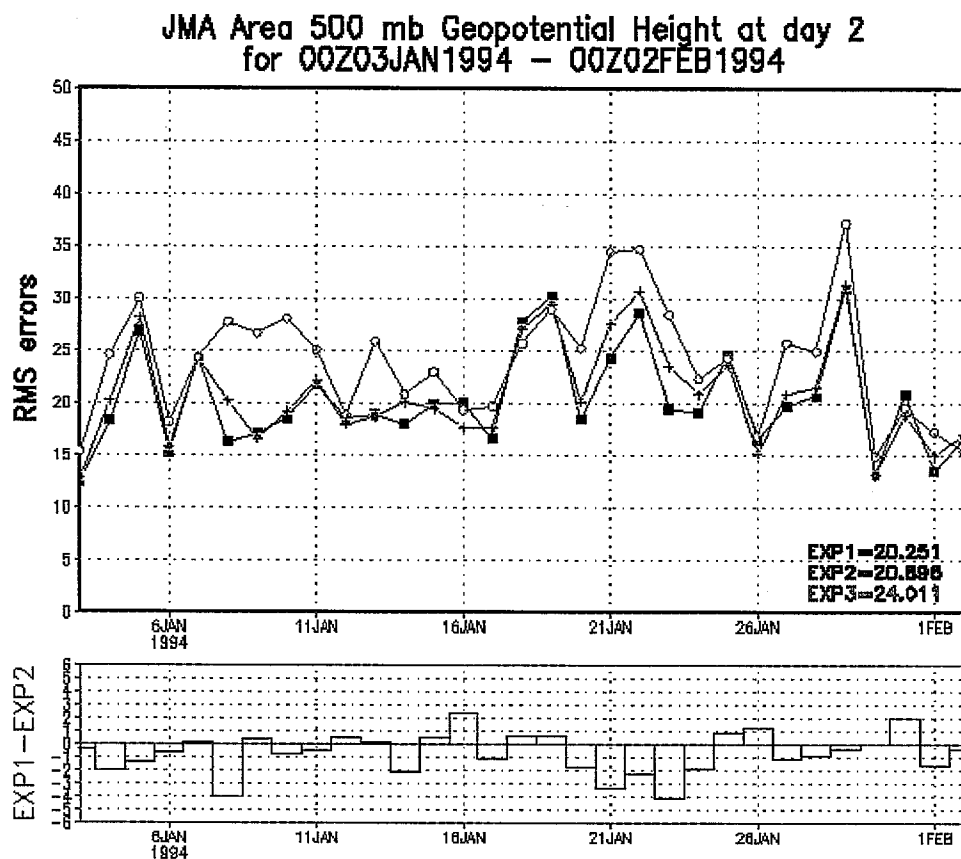


Figure 1

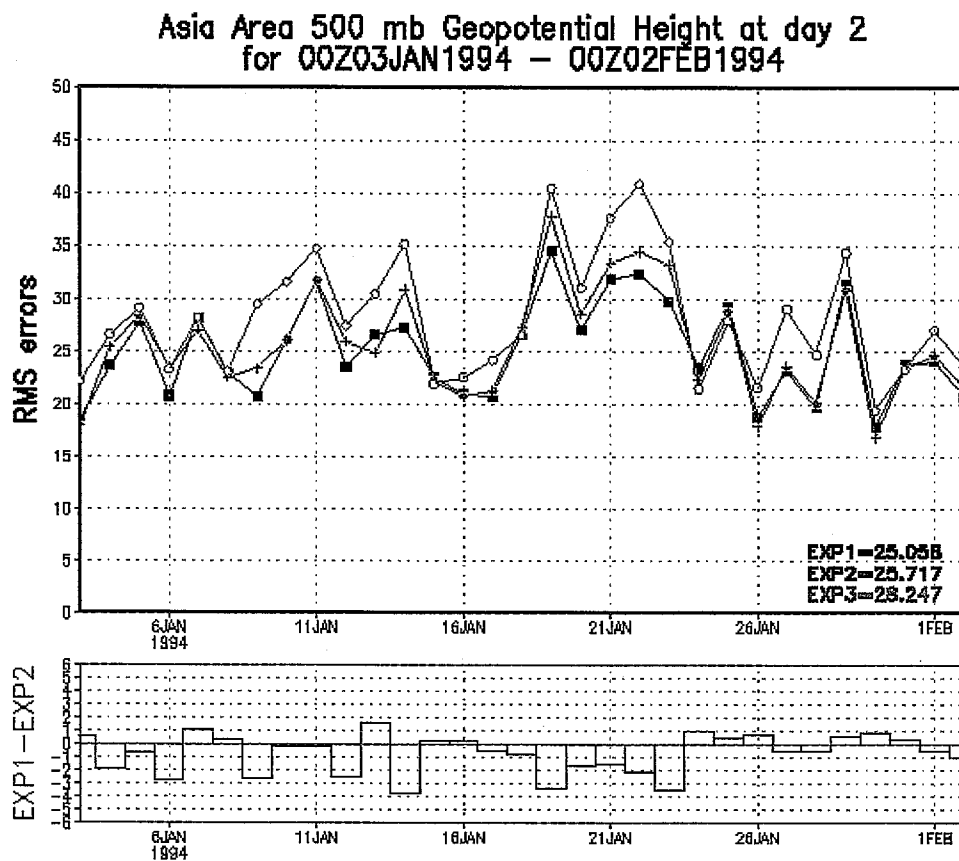


Figure 2

EXP 2 -1: 500Z DAY=2 FROM 012500

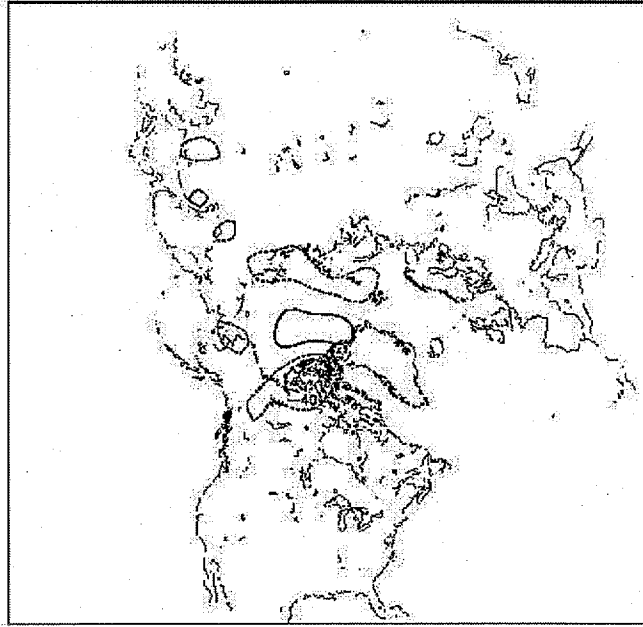
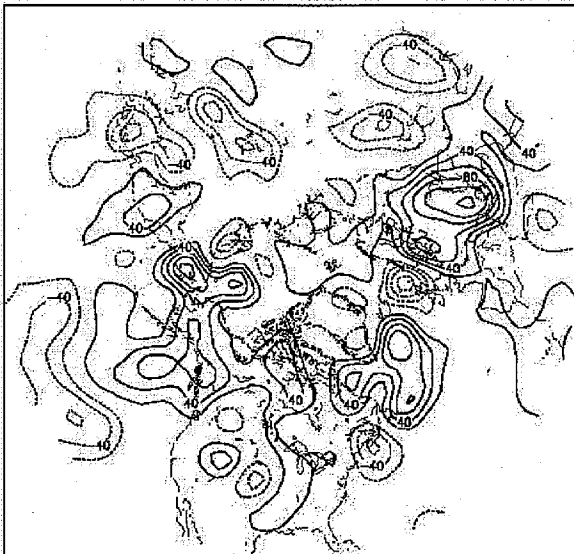


Figure 3. Day 2 EXP2 - EXP1; contour interval 20m.

VERIF - EXP 1: 500Z DAY=2 FROM 012500



VERIF - EXP 2: 500Z DAY=2 FROM 012500

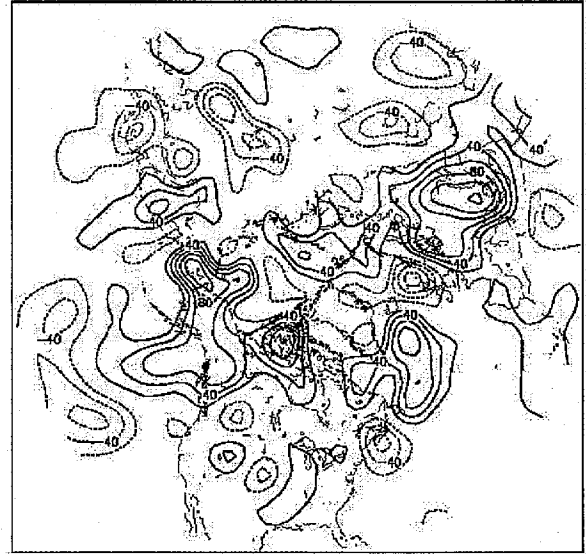


Figure 4. Day 2 forecast errors; EXP1 left, EXP2 right;  
Contour interval 20m.

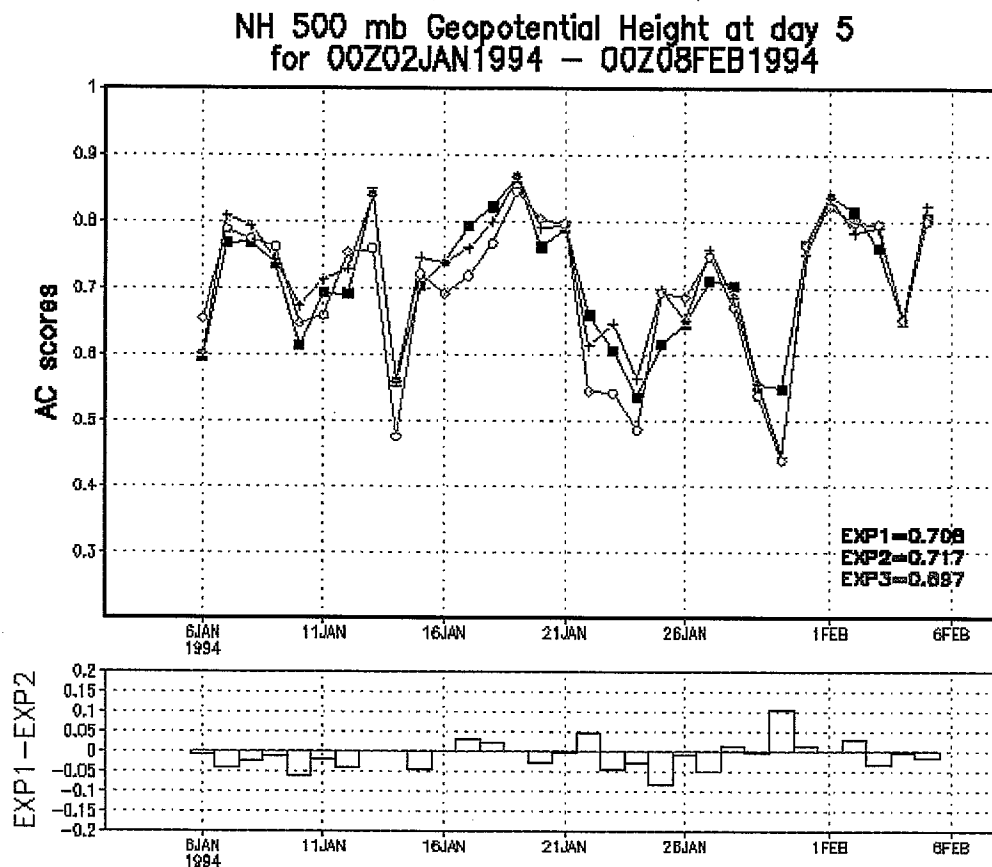


Figure 5.

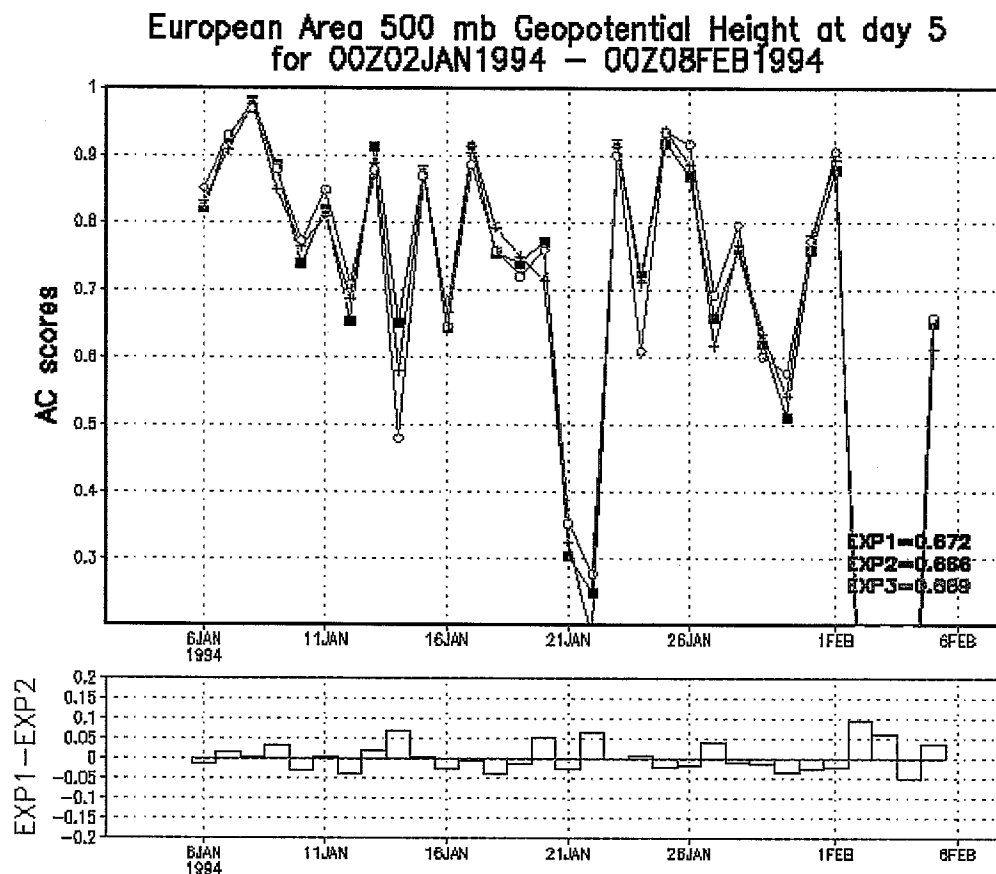


Figure 6

Asia Area 500 mb Geopotential Height at day 5  
for 00Z02JAN1994 – 00Z08FEB1994

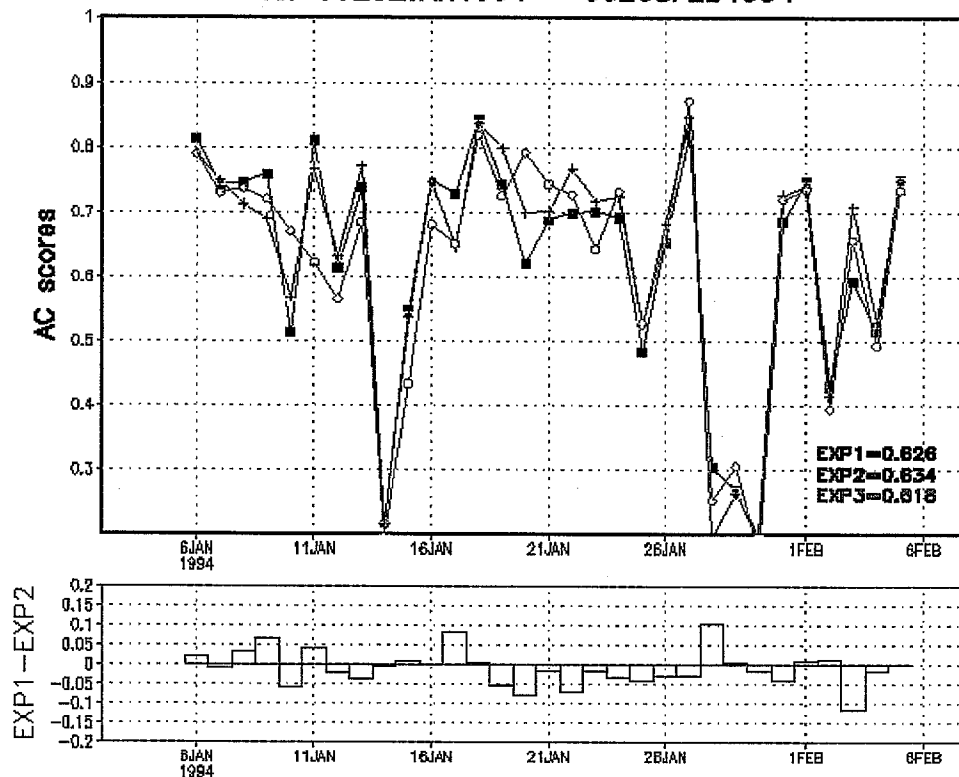


Figure 7

JMA Area 500 mb Geopotential Height at day 5  
for 00Z02JAN1994 – 00Z08FEB1994

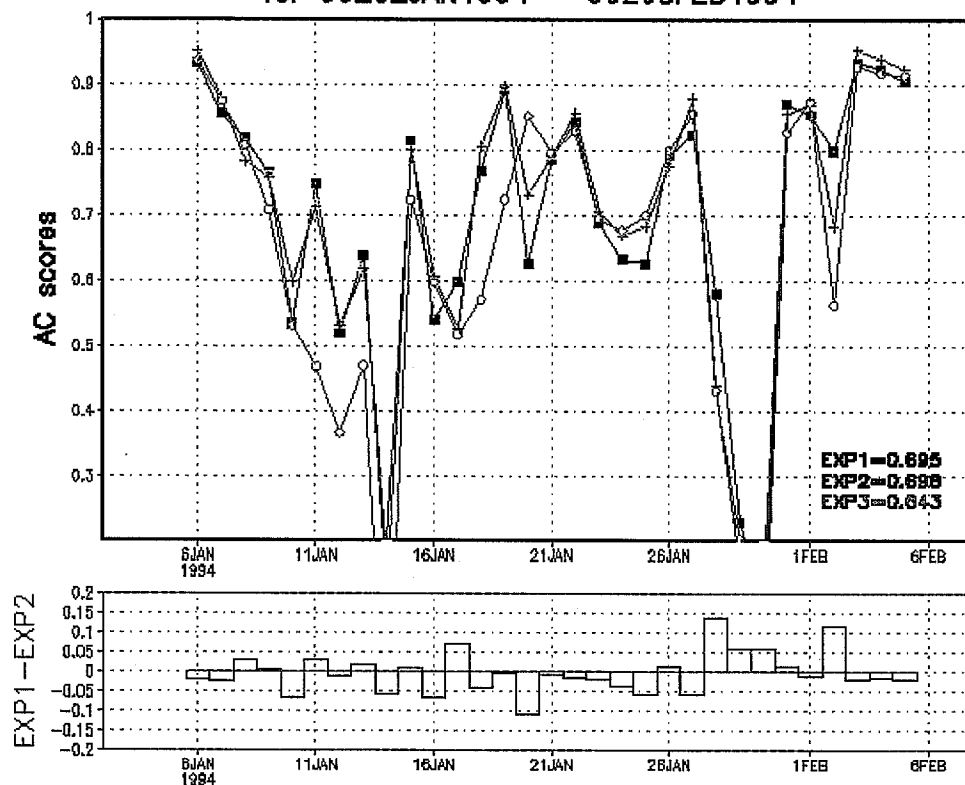
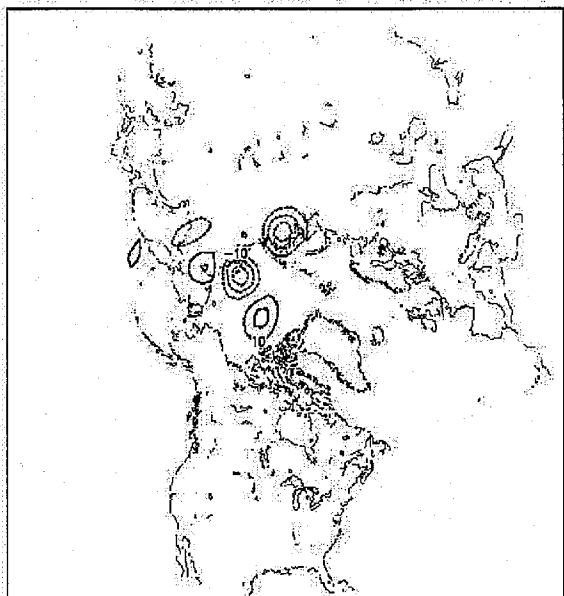


Figure 8

EXP 2 -1: 500Z DAY=0 FROM 012500



EXP 2 -1: 500Z DAY=5 FROM 012500

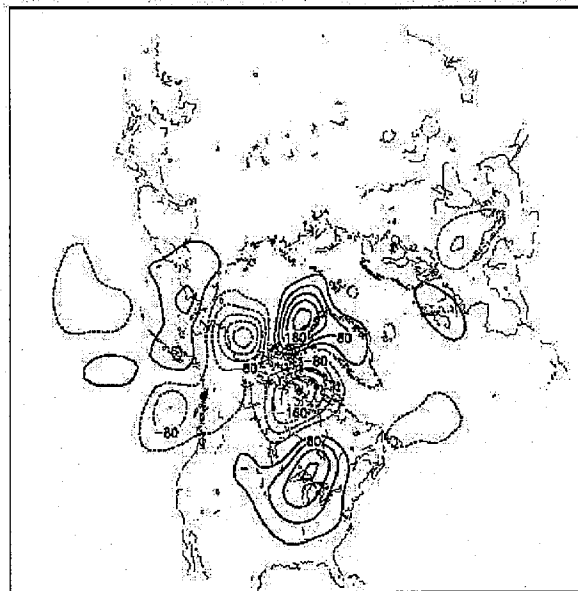
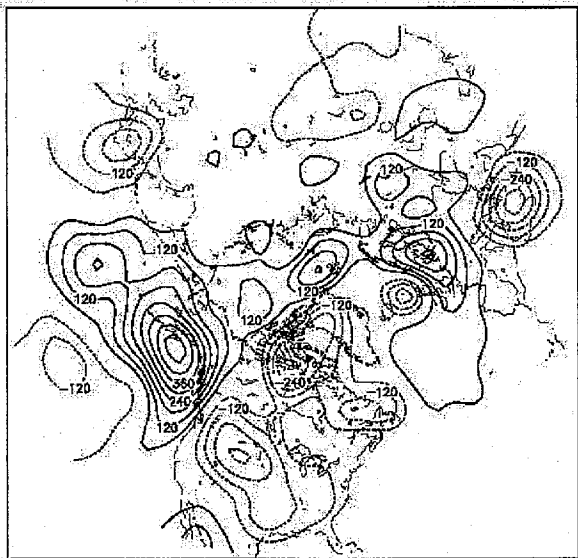


Figure 9. Day 0 (left) and day 5(right) EXP2-EXP1 differences. Contour interval 10m and 40m, respectively.

VERIF - EXP 1: 500Z DAY=5 FROM 012500



VERIF - EXP 2: 500Z DAY=5 FROM 012500

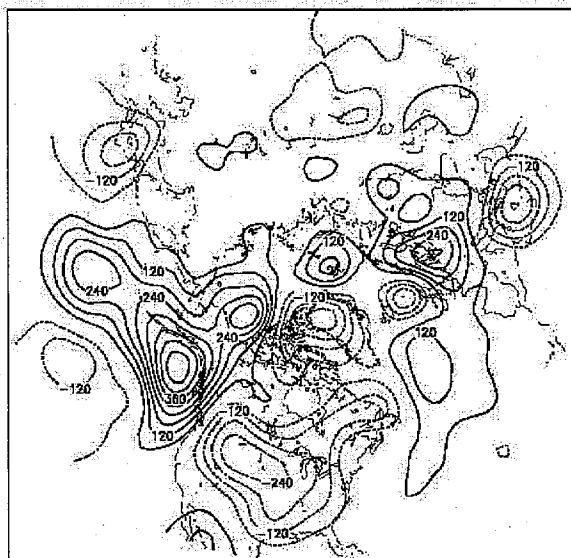


Figure 10. 5-day forecast errors; EXP1 left,EXP2 right. Contour interval 60m

U.S. DEPARTMENT OF COMMERCE  
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION  
NATIONAL WEATHER SERVICE  
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 432

NCEP/EMC OBSERVING SYSTEM EXPERIMENT (OSE) ON THE EFFECT OF  
LOSS OF RUSSIAN RAOB DATA

M. STEVEN TRACTON, ROBERT KISTLER AND YUEJIAN ZHU  
National Centers for Environmental Prediction/Environmental Modeling Center

AND

JIM GIRAYTYS  
WMO Contractor

MAY 2001

THIS IS AN UNREVIEWED MANUSCRIPT, PRIMARILY INTENDED FOR INFORMAL  
EXCHANGE OF INFORMATION AMONG NCEP STAFF MEMBERS.

# NCEP/EMC OBSERVING SYSTEM EXPERIMENT (OSE) ON THE EFFECT OF LOSS OF RUSSIAN RAOB DATA

by

M. Steven Tracton, Robert Kistler, and Yuejian Zhu  
National Centers for Environmental Prediction /Environmental Modeling Center  
NOAA/NWS, Washington, DC  
and  
Jim Giraytys  
WMO Contractor

## 1. Background

The number of sites in the Russian Federation taking upper air observations using radiosondes (RAOBs) decreased by 65% to 70% from January 1994 to December 1999, largely as a function of budget restrictions. Such a decrease might be expected to affect the skill of Numerical Weather Prediction forecasts. To address this issue, the Open Program Area Group on Data Processing and Forecast Systems of the WMO formed an Expert Team whose mission was to assess the possibility of using routine verification statistics as an alternative to conducting the more costly and time consuming Observing System Experiments (OSEs). The expert Team adopted the following working hypothesis:

*“It is possible to establish meaningful impact in the radiosonde network in Russia through evaluation of readily available verification scores of global and/or regional scale operational forecast models”.*

The results of this investigation are documented in the WMO publication, *“Study of the Impact of the Loss of Russian Federation RAOBS on NWP Verification Statistics in the Northern Hemisphere”* (available at <http://www.wmo.ch/web/www/reports.html>). In summary, those results neither proved nor disproved the hypothesis. No clear signal in the verification statistics was found, except *possibly* some degradation in skill over Asia and North America. Given changes in NWP systems over the years and likely natural variations in predictability (e.g., associated with circulation regime, seasonal trends), it is very difficult to detect a real signal and equally difficult to assign any change to a particular cause. One cannot conclude from this study that the loss of Russian Federation RAOBs does result in a meaningful loss of skill in NWP, only that the approach used in the investigation was not adequate to uncover a signal in the noise of the several factors which might lead to a change in routine verification scores.

To investigate further whether loss of the Russian data does have a meaningful impact on NWP skill, NCEP agreed to conduct an OSE using its Reanalysis system and data set. A brief description of the experiments is provided in Section 2. Illustration of results is provided in

Section 3, and a summary and discussion concludes the report (Section 4).

As will be shown, there is some systematic loss of skill in forecasts regionally (Asia, Alaska and Northern Canada) at short ranges (<3days). In the medium range (3-8 days) results are entirely consistent with the earlier investigation - no clear signal of a degradation in NWP forecast skill resulting from loss of the Russian RAOBs. In the OSE loss of the data unequivocally does produce an impact on forecasts, but beyond a few days that impact is largely random (some positive, some negative) and small relative to the total error in forecasts. The most important caveat to these results is that they apply only to the particular NWP system used and period of study selected.

## 2. Experimental Design

The OSE was run using the NCEP Reanalysis data assimilation system and data set. The system includes the NCEP global spectral model operational ("MRF") in 1995 with T62 (~210 km) horizontal resolution and a three-dimensional variational (3DVAR) analysis scheme. The data set includes RAOBs, TOVS temperature soundings, cloud-tracked winds, aircraft observations, land and ocean surface reports, etc. Further details can be found in Kistler, et al. (February, 2001 BAMS). The experiment period selected was January, 1994, a month where subjectively (based on appraisal of circulation patterns and weather systems) it appeared loss of the RUSSIAN RAOBs could have a notable impact upon NWP forecasts.

Three experiments were run:

EXP1: the control - all available observations

EXP 2: all observations, **except** those Russian RAOBS available in January, 1994, but NOT available in January, 2000 (nearly identical to the comparison between January, 1994 and December 1999 shown in Figs. 1 and 2, respectively, in the WMO report)

EXP 3: all RUSSIAN RAOBS removed

EXP3 was run as a baseline sensitivity test to assess the effect of removing all Russian Federation radiosonde reports. The relevant data sets were assimilated starting Dec. 1, 1993 and continued through Jan. 31, 1994 (December provided a one month "spinup"). Forecasts to 8 days were run from the 00Z analyses of each experiment for each day of January, 1994. Verifications were in the form of anomaly correlation (AC) scores and RMSE of forecasts with respect to the control set of analyses over the Northern Hemisphere and selected subregions thereof. Subjective case study evaluation was also performed to confirm and complement these objective verifications. Also, the subjective appraisal suggests that additional objective verifications against radiosonde observations likely would not change the results.

### 3. Results

#### 3.1 Short Range Forecasts

Verification statistics indicate that at shorter ranges a small, but systematic negative impact is felt in the Asian and JMA (centered about Japan) domains. This is demonstrated by the RMSE verifications for 2-day forecasts shown in Fig. 1 and 2. In these and the following figures black is for EXP1, the red for EXP2, and green for EXP3. The means over all cases is displayed on the bottom right of the top figure. The verification date (not initial time) is shown on the abscissa. The bottom portion of the figure displays the time series of the difference between EXP1 and EXP2.

The E1 errors on average over the period (20.25/25.06 for JMA/Asia) are slightly smaller than for EXP2 (20.89/25.72). More importantly, as seen clearly in the plots of EXP1-EXP2 scores, the negative impact occurs in most cases. Meteorological significance from subjective evaluation, however, is dubious. Over Alaska and Northern Canada, where there are no separate objective verifications, it appears qualitatively that forecast differences are larger than over Asia and the JMA regions. At day 2, and to a limited extent at day 3, there is a tendency generally for the EXP2 to have larger errors than EXP1. As an example, Fig. 3 shows for a particular case the difference between EXP1 and EXP2 2-day 500 mb height forecasts. In comparison to the corresponding EXP1 and EXP2 error fields (Fig. 4), one can observe that areas of forecast difference are reflected mostly as larger errors in EXP2. Note too the emergence of errors exclusive of the impact of loss of Russian data. This is evident from the observation that in regions of EXP1/EXP2 differences the errors are generally larger than those differences and more clearly from the errors occurring outside the areas of EXP1/EXP2 differences.

Experiments with regional models would have to be performed to assess more thoroughly the significance of the loss of data on short term forecasting over regions in relatively close proximity to the differences in analyses due to the decline in Russian RAOBs. In that context, before serious non linearities develop and other sources of error become dominant, the sense of the analysis differences (presumed to be negative) is more likely to be felt. Finally here, note from EXP3 that the impact that would occur from removal of all Russian RAOBs is considerably larger, as seen objectively in Figs. 1 and 2 and observed also in subjective evaluation. This is commented upon further in Section 4.

#### 3.2 Medium Range Forecasts

Figure 5 shows the time series of AC scores of 5-day forecasts computed for the Northern Hemisphere north of 20°. It can be seen that there is a mix of positive and negative impacts with the difference in the mean between EXP1 and EXP2 very small (E2 actually better) and certainly not statistically significant. The same conclusion is drawn from the corresponding set of RMS errors (not shown). And, the same result applies to each of the verification sub regions shown in Figs. 6 -8 and for the most part to day 3 forecasts as well (not shown).

The lack of significant effect on verification scores does not mean that the loss of the Russian RAOBS does not have an impact on the forecasts. This can be seen from the initial and 5- day EXP2-EXP1 differences shown in Fig. 9 for the same case discussed above. The relatively small differences that exist between analyses (day 0) in the polar regions over and to the north of Siberia evolve (via translation and downstream amplification/propagation) through those shown in Fig. 3 (day 2) to seemingly very significant levels both in magnitude and areal coverage. But, beyond 2-3 days those differences are generally small compared to the forecast error, whether it

be for EXP1 or EXP2 (Fig. 10). In effect, the signal of forecast impact due to loss of the Russian RAOBs is essentially noise in the context of the total forecast error. Or in the vernacular of NWP, the forecasts are much more alike than either is to the real atmosphere. From one case to the next or within one case from region to region subjective appraisal shows that, when it is possible to judge that one forecast is better than the other in some respect, there is no systematic preference for EXP1 or EXP2. By no later than days 2- 3, non-linear interactions dominate and the sense of forecast differences is essentially random.

#### 4. Summary and Discussion

The OSE was designed explicitly to isolate the effect of the loss of Russian RAOBs. Unlike the earlier study based on evaluation of routine verification scores, the impact is not obscured by before and after differences in atmospheric predictability (e.g., related to circulation regime), changes in the global observing system, or differences between models and data assimilation schemes. It was a perfectly "clean", *but limited*, experiment.

The results indicated a small, but systematic loss of skill at short ranges (< 3days) in the regions most local to the areas affected by the loss of data (Asia, JMA, Alaska and Northern Canada). In the medium range (3-8 days), results were remarkably comparable to those arrived at by the Expert Team - no degradation in forecast skill as a result of deterioration in the Russian Federation RAOB network. Beyond day 3, the loss of data does have a notable impact, but the differences between the with and without experiments (EXP1 and EXP2) are small and not systematic relative to the total error of either EXP1 or EXP2 predictions. The loss in skill associated with loss of the RAOBs over Russia is essentially noise in the context of other sources of error, i.e., analysis errors exclusive of the Russian data and inadequacies in the data assimilation system and forecast model.

It is worthwhile to note that the objective verifications and subjective evaluation of EXP3 does indicate a fairly large and more or less systematic degradation in forecast skill even at medium ranges as a result of removing all Russian Federation RAOBS. In most instances the result is making a bad forecast even worse. But in some, the loss of data clearly renders a reasonably good prediction less useful. In the context of this experiment, by chance or design, it appears the decrease, but not elimination, in Russian RAOBS was such as to minimize the impact on global NWP.

Of course, the principal caveat of this OSE is that it applies only to the particular numerical and analysis forecast system used (NCEP, 1995 vintage) and only for the particular period selected. To adequately generalize would require using the latest state-of-the art system (more than one) applied to several independent sample periods. That would be exceedingly costly in human and computer resources (and beyond anything currently planned at NCEP). And, if later studies, in fact, did demonstrate a clear and meteorologically significant loss in skill from the decline in Russian RAOBs, one likely would have to address the cost effectiveness of alternative observing systems (e.g., aircraft ascent/decent soundings) in comparison to reconstituting the Russian RAOB network.

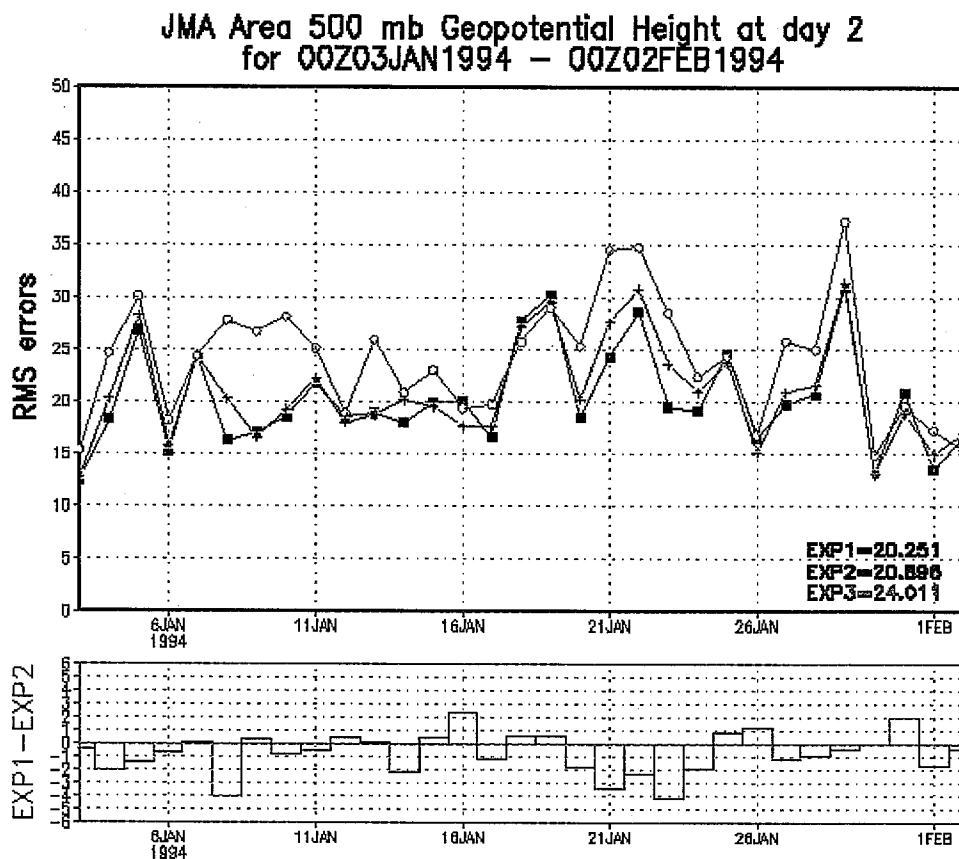


Figure 1

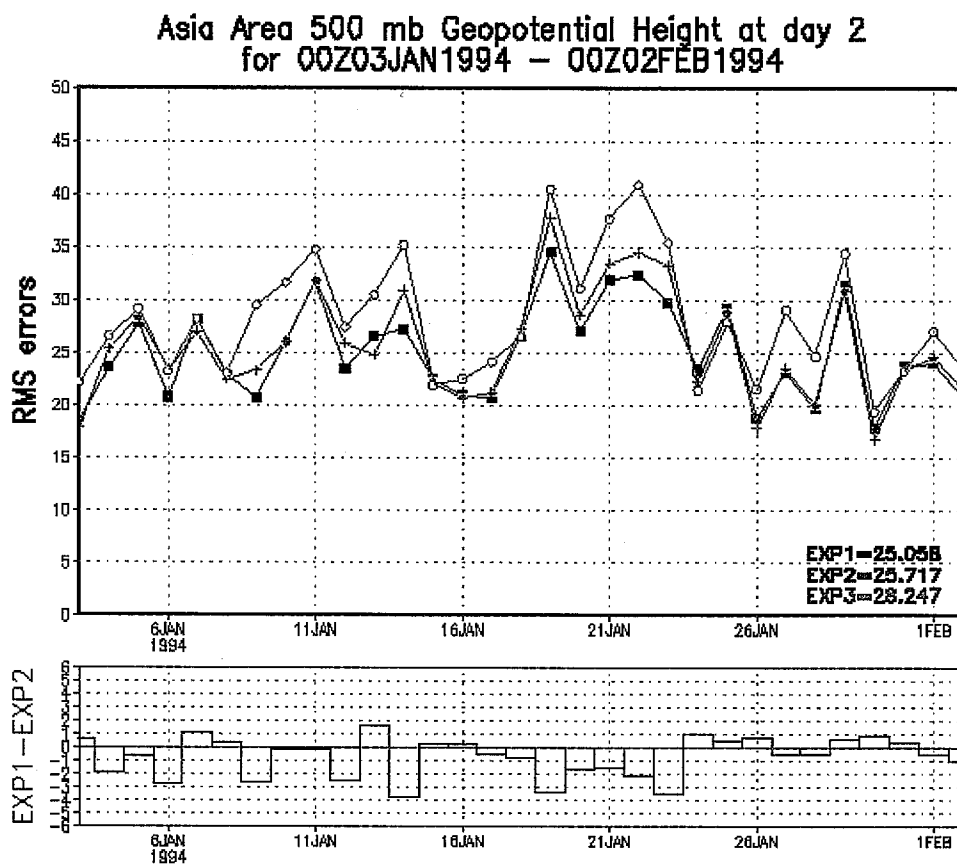


Figure 2

EXP 2 -1; 500Z DAY=2 FROM 012500

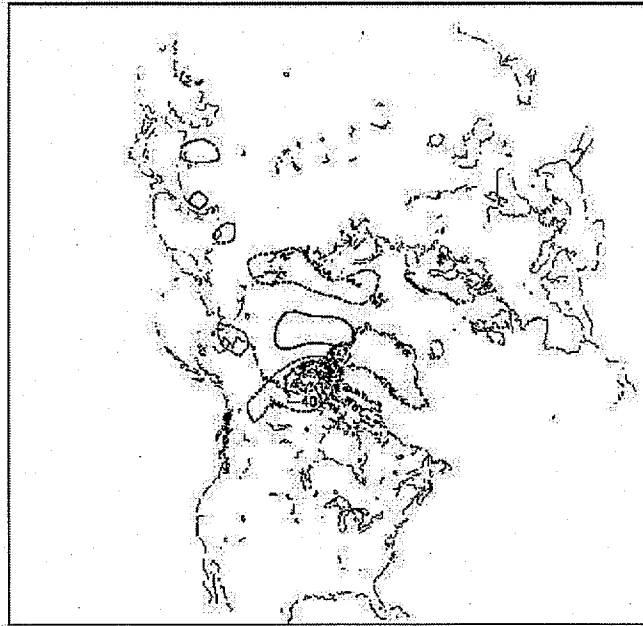
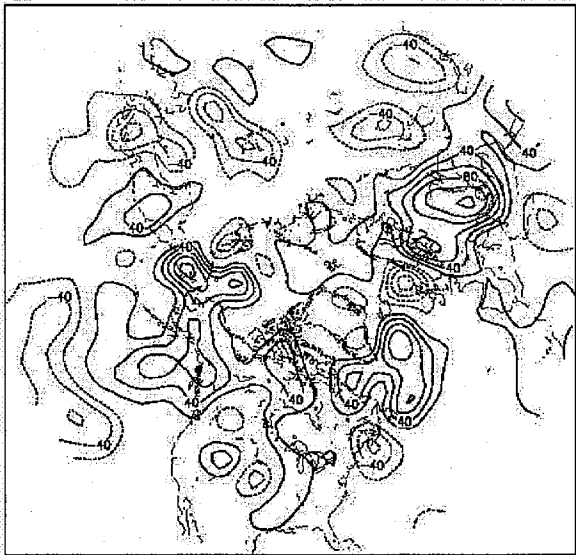


Figure 3. Day 2 EXP2 - EXP1; contour interval 20m.

VERIF - EXP 1; 500Z DAY=2 FROM 012500



VERIF - EXP 2; 500Z DAY=2 FROM 012500

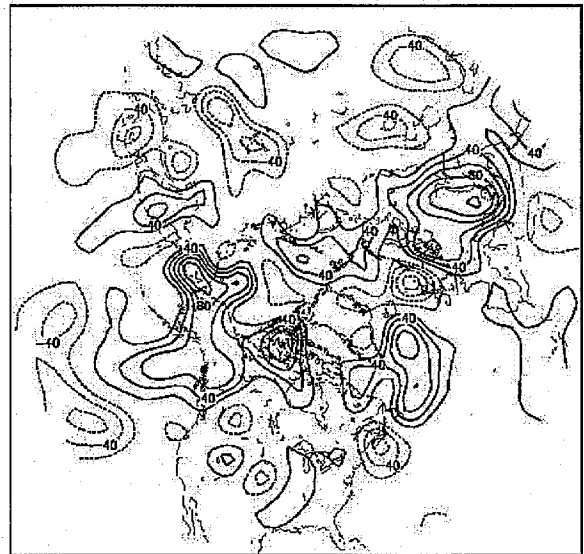


Figure 4. Day 2 forecast errors; EXP1 left, EXP2 right;  
Contour interval 20m.

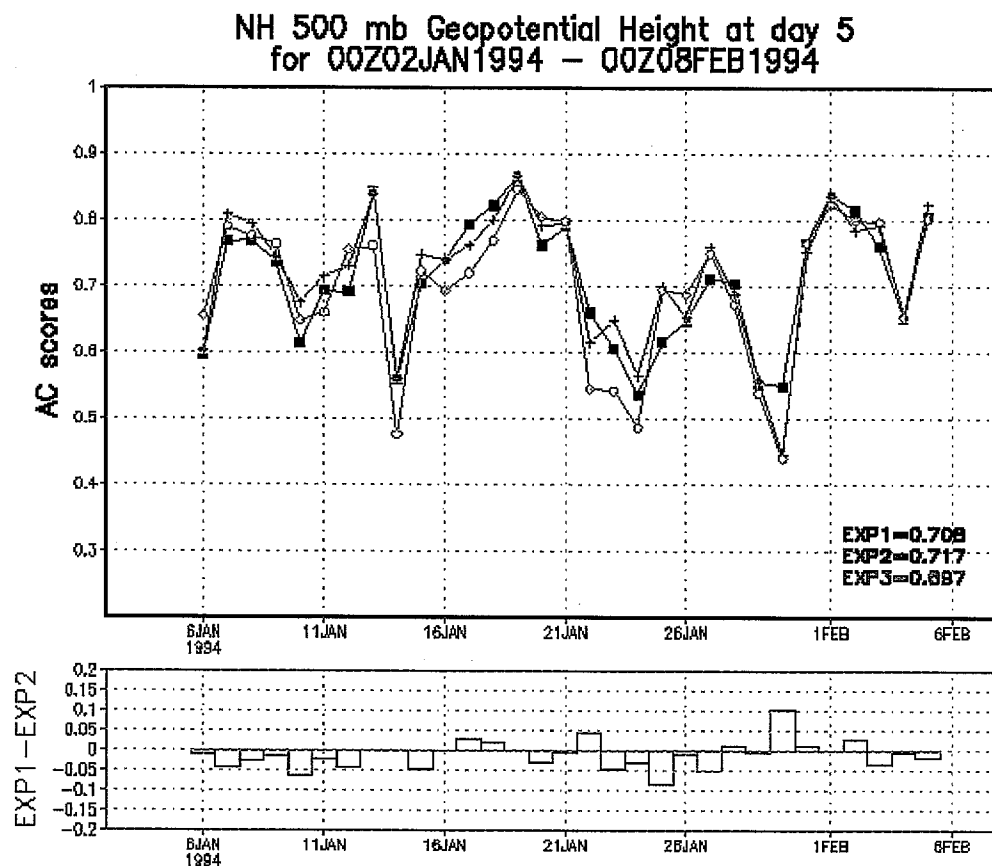


Figure 5.

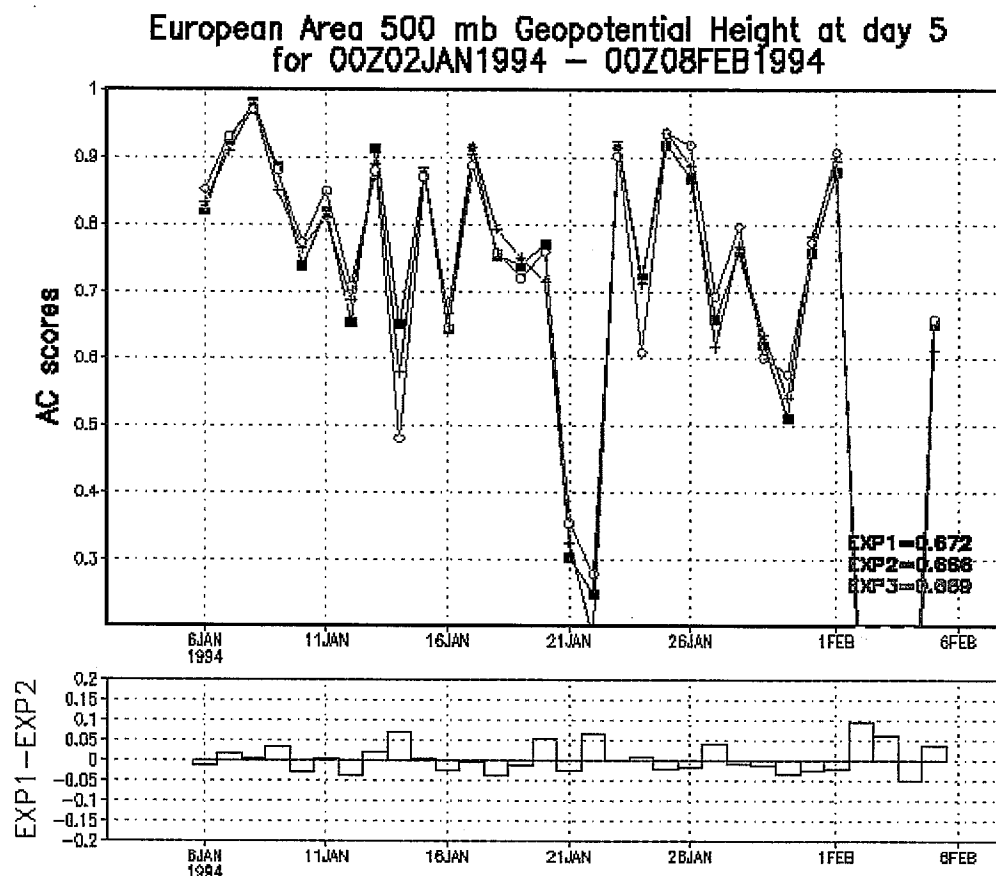


Figure 6

Asia Area 500 mb Geopotential Height at day 5  
for 00Z02JAN1994 – 00Z08FEB1994

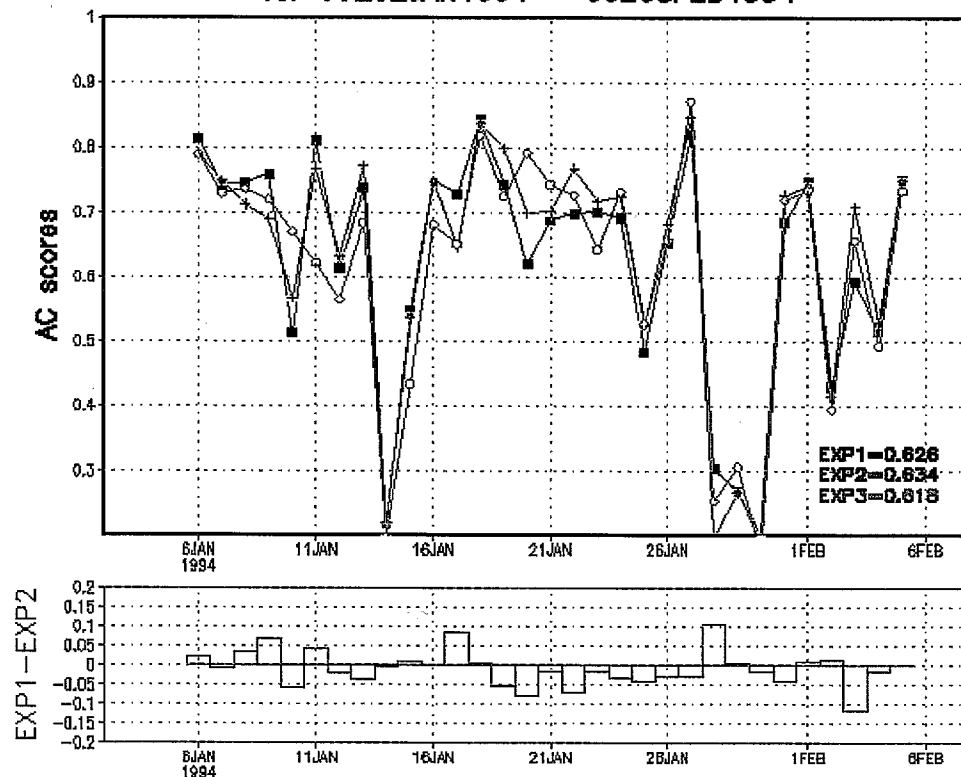


Figure 7

JMA Area 500 mb Geopotential Height at day 5  
for 00Z02JAN1994 – 00Z08FEB1994

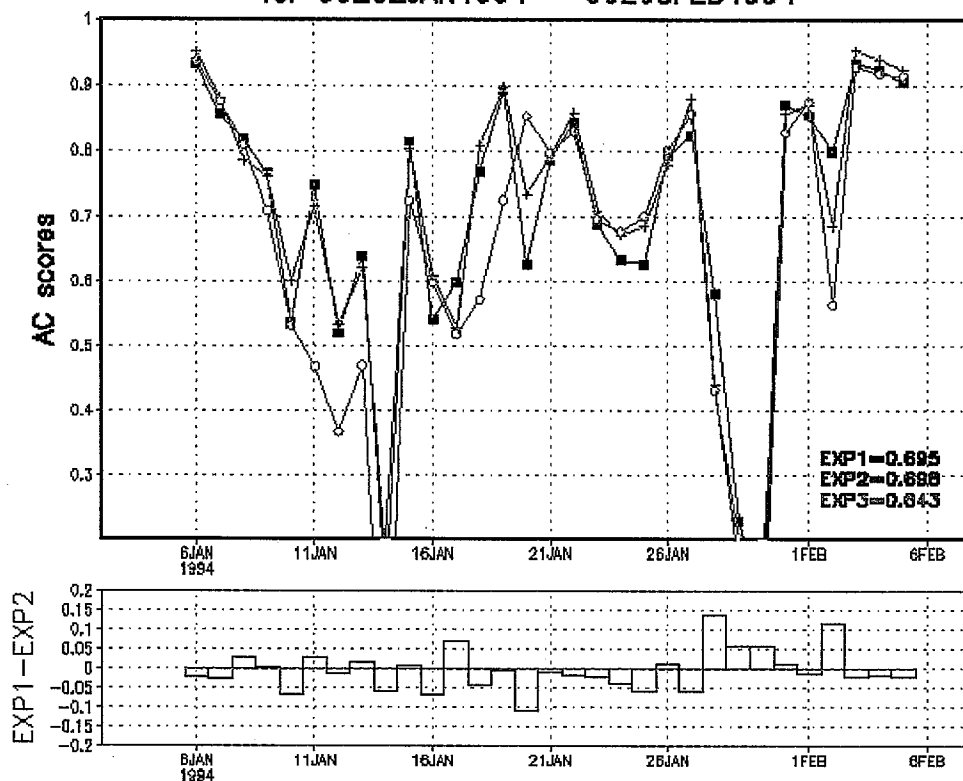
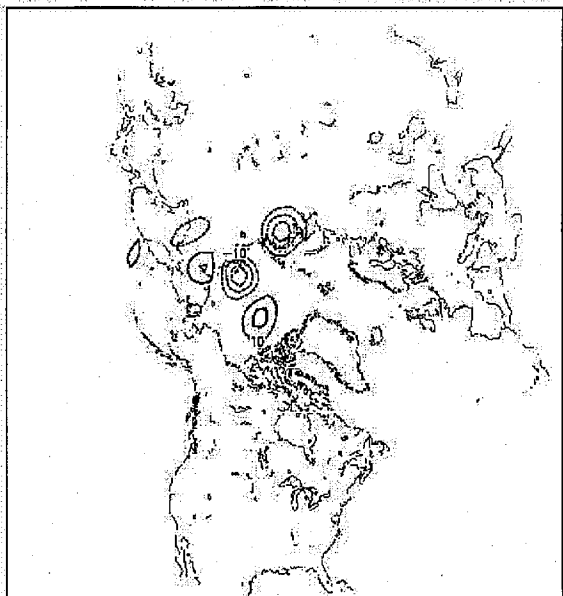


Figure 8

EXP 2 -1: 500Z DAY=0 FROM 012500



EXP 2 -1: 500Z DAY=5 FROM 012500

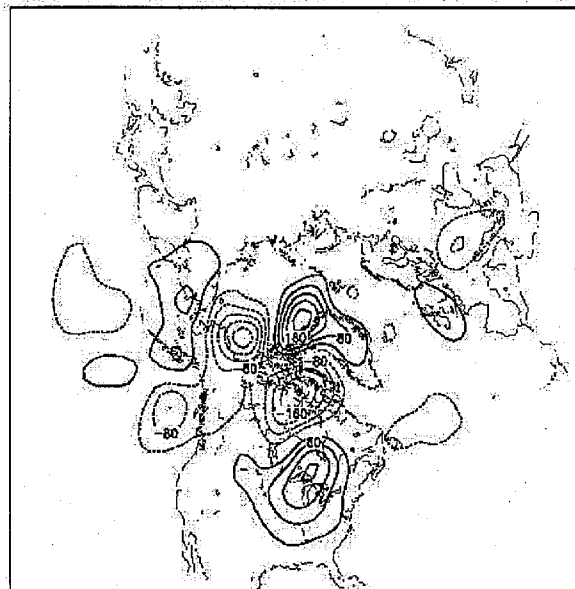
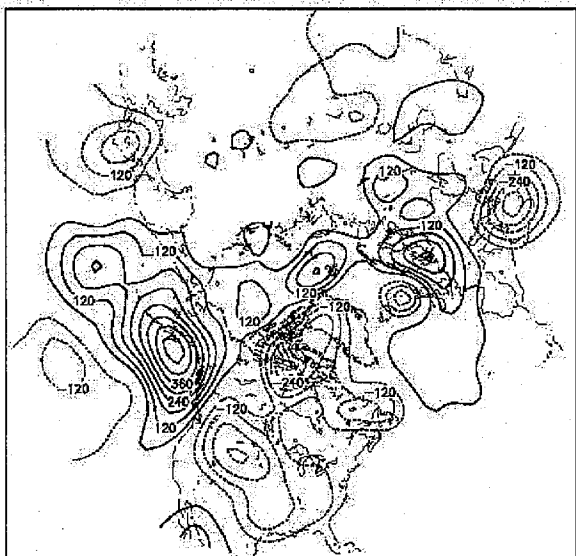


Figure 9. Day 0 (left) and day 5(right) EXP2-EXP1 differences. Contour interval 10m and 40m, respectively.

VERIF - EXP 1: 500Z DAY=5 FROM 012500



VERIF - EXP 2: 500Z DAY=5 FROM 012500

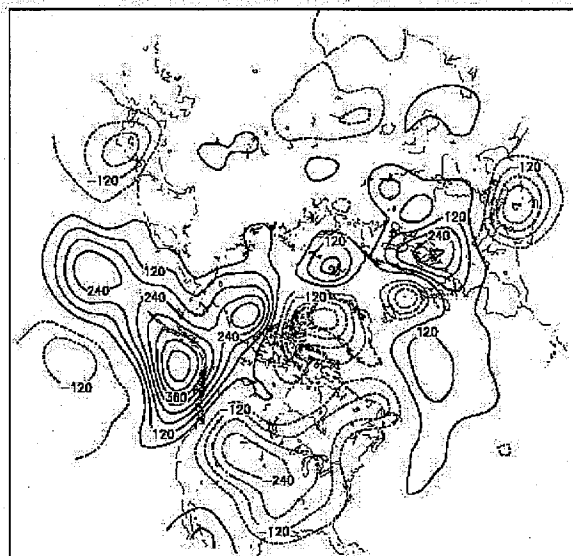


Figure 10. 5-day forecast errors; EXP1 left,EXP2 right. Contour interval 60m